The AAPS Journal 2004; 6 (3) Article 19 (http://www.aapsj.org).

The Back-Step Method – Method for Obtaining Unbiased Population Parameter Estimates for Ordered Categorical Data

Submitted: October 6, 2003; Accepted: February 20, 2004; Published: August 11, 2004.

Maria C. Kjellsson,[1] Siv Jönsson,[1] and Mats O. Karlsson[1]

[1]Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden

ABSTRACT

A significant bias in parameters, estimated with the proportional odds model using the software NONMEM, has been reported. Typically, this bias occurs with ordered categorical data, when most of the observations are found at one extreme of the possible outcomes. The aim of this study was to assess, through simulations, the performance of the Back-Step Method (BSM), a novel approach for obtaining unbiased estimates when the standard approach provides biased estimates. BSM is an iterative method involving sequential simulation-estimation steps. BSM was compared with the standard approach in the analysis of a 4-category ordered variable using the Laplacian method in NONMEM. The bias in parameter estimates and the accuracy of model predictions were determined for the 2 methods on 3 conditions: (1) a nonskewed distribution of the response with low interindividual variability (IIV), (2) a skewed distribution with low IIV, and (3) a skewed distribution with high IIV. An increase in bias with increasing skewness and IIV was shown in parameters estimated using the standard approach in NONMEM. BSM performed without appreciable bias in the estimates under the 3 conditions, and the model predictions were in good agreement with the original data. Each BSM estimation represents a random sample of the population; hence, repeating the BSM estimation reduces the imprecision of the parameter estimates. The BSM is an accurate estimation method when the standard modeling approach in NONMEM gives biased estimates.

KEYWORDS: ordered categorical, proportional odds model, bias in parameter estimates, NONMEM, Laplacian, pharmacodynamics.

INTRODUCTION

The nonlinear mixed effects modeling software, NONMEM,[1] may in some situations provide parameter estimates with an apparent bias. It has previously been shown that both the first order (FO) method and the first order conditional estimation (FOCE) method in NONMEM can produce biased pharmacokinetic (PK) parameter estimates if the variability of the data are high.[2,3] The FOCE method has also been shown to produce considerable bias in pharmacodynamic (PD) parameters.[4] More recently, Jonsson et al[5] showed that the Laplacian method in NONMEM produces biased parameter estimates when using a nonlinear pharmacodynamic model to describe the data. In these situations, the goodness-of-fit and the predictive performance of the model may improve with an estimation method resulting in unbiased parameter estimates.

Ordered categorical data are commonly used to describe subjectively scored symptoms and side effects and most of the observations are often at one extreme of the possible outcomes (ie, the distribution of response is skewed). The standard approach for modeling ordered categorical data is the logit model for cumulative probabilities, also referred to as the proportional odds model. Repeated measurements are often handled using mixed effects modeling, where the interindividual variability (IIV) is added as an overall variability on baseline probabilities.[6-9] When analyzing ordered categorical data with a skewed distribution using the standard mixed effects modeling approach, the parameter estimates will be severely biased if the Laplacian method in NONMEM is used.[10] The bias in the parameter estimates increases with increasing skewness of the response distribution and increasing IIV of data. The frequency of rare events will be overestimated when simulation of new data is performed using the biased parameters.

One way of addressing the problem is to normalize the distribution of the categorical outcome by redefining the categories. If the categories were chosen before the data was collected, or the data was categorized on collection, redefining the categories would imply recollection of data using the new definitions for the categories. In these cases, it might be preferable to reduce bias in parameter estimates by changing the model. Two methods have been proposed that both view the modeling of categorical responses as a 2-step process[11,12] in which the first step is to model the incidence of response or the probability of being a responder and the second step is to model the severity of response.

Corresponding Author: Maria C. Kjellsson, Division of Pharmacokinetics and Drug Therapy, Department of Pharmaceutical Biosciences, Uppsala University, Box 591, SE-751 24, Uppsala, Sweden. Tel: +46 18 471 4418. Fax: +46 18 471 4003. Email: maria.kjellsson@farmbio.uu.se.

Olsen and Schafer[11] propose to model the response as a semicontinuous variable, using a logistic model for modeling the incidence of effect, while modeling the severity of effect using a continuous model, both steps including IIV. In the first step, data are reduced from polychotomous to dichotomous by modeling the incidence of effect/no effect (not considering the severity of the effect), thereby decreasing the skewness of response and, consequently, reducing the bias in parameter estimates.

In the modeling procedure proposed by Kowalski et al,[12] the incidence of being a responder/non-responder is modeled in the first step, considering only one observation per individual and no IIV. In the second step, the severity of response (given being a responder) is modeled using the standard mixed effects modeling approach. Bias in parameter estimates is avoided by not considering IIV in the first step and, similar to Olsen and Schafer's model,[11] by reducing the skewness of the response distribution by only including responders in the second step. However, since the first step only considers one observation per patient, dose escalation studies and time-varying covariates cannot be appropriately handled using this method. Another limitation is that in the first step the parameter estimates will be dependent on the number of observations per subject, with increasing probability of a nonzero event with increasing number of observations. Such data-dependency of parameters limits the possibility of making extrapolations based on the model.

An alternative to changing the model to handle biased estimates is to change or modify the estimation procedure itself. The Back-Step Method (BSM) is an iterative method that searches for the unbiased parameter estimates, which upon simulation generate data that mimic the original data. The purpose of this investigation is to assess the performance of the BSM through simulation and estimation of ordered categorical data.

MATERIALS AND METHODS

This is a Monte Carlo simulation study in which the original data sets were derived from a known model. All simulations and model fittings were performed using NONMEM Version VI (beta).[1] No differences were found for selected data sets when parameter estimates of NONMEM Version VI (beta) and NONMEM Version V were compared, supporting that the BSM could be used with NONMEM Version V. The Laplacian estimation method with the likelihood option was used.

In the following section, the proportional odds model is first outlined, followed by a description of the simulation conditions, the BSM, the methods for calculation of bias and imprecision contribution, and the method for obtaining standard errors for BSM estimates.

The bias of the standard approach in NONMEM has previously been reported by the authors,[10] but results are also given here for comparison with the BSM.

The Proportional Odds Model

The severity of response was assessed on a 4-category ordinal scale: 0, 1, 2, and 3. A proportional odds model, similar to the models used by Gupta et al[6] and Sheiner et al[9] was used for simulation and estimation of the probability of events. If $Y_i = (Y_{i1}, Y_{i2}, \ldots Y_{in})$ is the vector of categorical response for the ith individual, then the probability that Y_{it} is greater than or equal to the score m (m = 0, 1, 2, and 3) has the following general structure:

$$f\left[P\left(Y_{\mathrm{it}} \geq m \middle| \eta_{\mathrm{i}}\right)\right] = \mathrm{logit}\left(p_{\mathrm{i}}\right) = f_{\mathrm{m,i}} + \eta_{\mathrm{i}} \qquad (1)$$

in which,

$$\mathrm{logit}\left(p_{\mathrm{i}}\right) = \log\left[\frac{p_{\mathrm{i}}}{1-p_{\mathrm{i}}}\right] \qquad (2)$$

giving,

$$P\left(Y_{\mathrm{it}} \geq m \middle| \eta_{\mathrm{i}}\right) = \frac{e^{f_{\mathrm{m,i}}+\eta_{\mathrm{i}}}}{\left(1+e^{f_{\mathrm{m,i}}+\eta_{\mathrm{i}}}\right)} \qquad (3)$$

The function $f[x]$ denotes the logit transform of a probability, which was used to ensure the probability to be between zero and one. The $f_{\mathrm{m,i}}$ is a function of baseline conditions, drug dose, and placebo as defined in Equation 4, and η_{i} is a normally distributed, zero mean random variable with standard deviation ω describing interindividual variability.

$$f_{\mathrm{m,i}} = \sum_{j=1}^{m} \theta_{\mathrm{bl=j}} + \theta_{\mathrm{plc}} \cdot I_{\mathrm{plc}} + \theta_{\mathrm{dose}} \cdot Dose \qquad (4)$$

in which $\theta_{\mathrm{bl=j}}$ specifies the baseline probabilities of the different levels of Y. The intercept parameter $\theta_{\mathrm{bl=0}}$ (for m = 0) need not be estimated because the cumulative probability of event score being zero or more is one. I_{plc} is an indicator variable for placebo effect taking the value zero at the first occasion and the value one at all other occasions. The θ_{plc} and θ_{dose} are fixed effect parameters describing the magnitude of the placebo effect and dose effect, respectively.

All parameters were vectorized in the vector Φ to ease notation:

$$\mathbf{\Phi} = \left[\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6\right] = \left[\theta_{\mathrm{bl=1}}, \theta_{\mathrm{bl=2}}, \theta_{\mathrm{bl=3}}, \theta_{\mathrm{plc}}, \theta_{\mathrm{dose}}, \omega^2\right] \qquad (5)$$

The first 5 elements of the vector are the fixed effect parameters and the last element is the variance of the random effect.

Table 1. Nominal Parameter Values and Expected Proportions of Observations to Each Category in the Simulated Original Data Sets, Presented as Percent of the Total Population*

Condition	Type of Distribution	Expected Proportion of Observations to Category 0/1/2/3 at Baseline (%)	Nominal Parameter Values θ_1	θ_2	θ_3	θ_4	θ_5	ω^2
(1)	nonskewed	24 / 26 / 26 / 24	1.85	-1.85	-1.85	0.483	0.0459	4
(2)	skewed	96.5 / 1.22 / 1.44 / 0.84	-4.88	-0.548	-1.18	1.55	0.0303	4
(3)	skewed	96.5 / 1.22 / 1.44 / 0.84	-11.8	-1.32	-2.96	3.85	0.0717	40

* Nominal parameter values were set to simulate 3 conditions: (1) nonskewed distribution of response with low interindividual variability (IIV), (2) skewed distribution with low IIV, and (3) skewed distribution with high IIV.

Indexes to Φ were used to indicate parameter estimates from the standard mixed effects modeling approach (STD), the simulation step in the BSM procedure (sim), the estimation step in the BSM procedure (est), and the final parameters estimated using BSM (BSM).

By fitting above described proportional odds model to data, the parameter estimates from the standard mixed effects modeling approach, Φ_{STD}, were obtained.

Simulation Conditions

The data sets consisted of 1000 patients, evenly divided into 4 dosage groups (placebo, 7.5, 15, and 30 units of drug), with 4 observations per individual: 1 baseline observation and 3 after study drug intervention.

Nominal parameter values were set to simulate 3 conditions; (1) a population with a nonskewed distribution of the response and low IIV, (2) a skewed distribution of the response and low IIV, and (3) a skewed distribution of the response and high IIV (Table 1). In the population with the nonskewed distribution, nominal parameter values were chosen such that, at baseline, observations of each category were equally frequent. The expected proportion of observations, to be in one category, is the same for the 2 skewed distributions, but the nominal parameter values used to generate these data differ, due to differences in variability. The skewed distribution with high IIV was designed to mimic a real drug data set of observations of an adverse event.

The Back-Step Method

The Back-Step Method is described below and shown in Figure 1. The BSM searches for the unbiased parameter estimates, Φ_{BSM}, which upon simulation generate data that mimic the original data. This search is an iterative process involving sequential simulation-estimation steps.

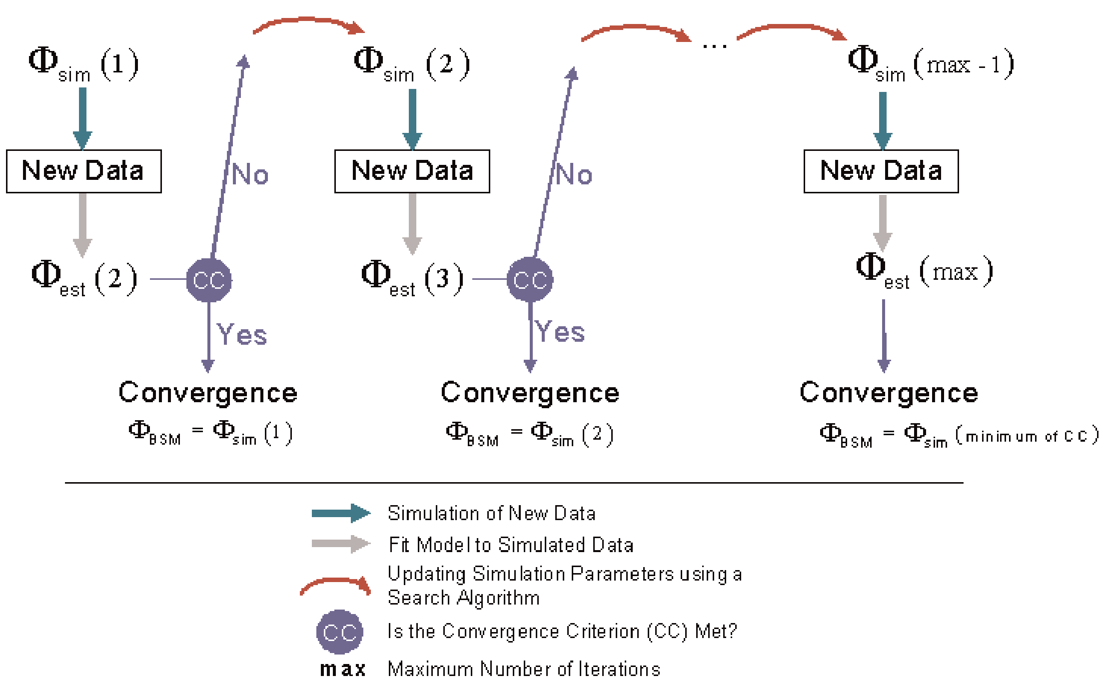

Figure 1. The Back-Step Method. Iteration of sequential simulation-estimation, followed by updating simulation parameters is repeated until the CC is met or the maximum number of iterations is exceeded. In the case of process ending due to maximum number of iterations reached, the Φ_{sim} with the lowest value of CC is kept as the final parameter estimates, Φ_{BSM}. Φ_{sim} is the vector of simulation parameters and Φ_{est} is the vector of parameters estimated from simulated data.

The proportional odds model was fitted to the original data, generating the first set of estimated parameters, $\Phi_{est}(1)$. These parameters were used as the first simulation parameters, $\Phi_{sim}(1)$, when simulating new data under the model. The new data were fitted to the model and the second set of estimated parameters, $\Phi_{est}(2)$, was generated. Based on the deviation between $\Phi_{est}(1)$ and $\Phi_{est}(2)$, the previous simulation parameters, $\Phi_{sim}(1)$, were updated according to the search algorithm, Equation 6, giving the new simulation parameters, $\Phi_{sim}(2)$:

$$\phi_{l,sim}(n+1) = \phi_{l,sim}(n) \cdot \left(1 + \frac{\left(\phi_{l,est}(1) - \phi_{l,est}(n)\right)}{\phi_{l,est}(n)}\right) \quad (6)$$

where $\phi_{l,\text{sim}}(n+1)$ and $\phi_{l,\text{sim}}(n)$ are simulation parameters for the (n + 1)th and the nth iteration, respectively; $\phi_{l,\text{est}}(n)$ and $\phi_{l,\text{est}}(1)$ are the estimated parameters from the nth and the first iteration, respectively. The calculations were performed individually for each of the parameters, as indicated by the parameter index *l* of the vectors. Each simulation-estimation step was followed by an update of the simulation parameters based on Equation 6. This cycle was repeated until the maximum number of iterations (350) was exceeded or the convergence criterion (CC) was less than one. CC is given by the following:

$$\sum_{l=1}^{\text{no. of parameters}} \frac{\left|\phi_{l,\text{est}}(1) - \phi_{l,\text{est}}(n)\right|}{\text{SE}\left(\phi_{l,\text{est}}(1)\right)} \quad (7)$$

where $\phi_{l,\text{est}}(n)$ and $\phi_{l,\text{est}}(1)$ are the estimated parameters for the nth and first iteration and $\text{SE}(\phi_{l,\text{est}}(1))$ is the standard error of the first estimated parameter. If the BSM was ended, because the CC was less than one, the Φ_{sim} of that iteration was kept as the final parameter estimate. On the other hand, if the BSM was ended because the maximum number of iterations was exceeded, the Φ_{sim} of the previous iteration with the lowest value of the CC was kept as the final parameter estimate, Φ_{BSM}. The above procedure was automated using the programming language Perl,[13] and the code is available on request from the corresponding author.

Estimation of Bias

To assess the bias in the parameter estimates, 100 original data sets were simulated for each of the conditions (1), (2), and (3) using the nominal parameter values listed in Table 1. Each original data set was then analyzed once with the standard mixed effects modeling approach, resulting in one Φ_{STD} for each simulated data set, and once with the Back-Step Method, resulting in one Φ_{BSM} for each simulated data set. The nominal parameters used for simulating the original data sets were taken to be the true unbiased parameter value, Φ. The relative biases were calculated according to Equation 7, for each parameter individually, as indicated by *l*, the parameter index of the vectors.

$$\text{bias}_{\text{BSM}} = \frac{\phi_{l,\text{BSM}} - \phi_l}{|\phi_l|}$$
$$\text{bias}_{\text{STD}} = \frac{\phi_{l,\text{STD}} - \phi_l}{|\phi_l|} \quad (8)$$

Estimation of Contribution of Imprecision

Imprecision in the parameter estimates from the BSM procedure will arise from 2 sources. The first source of imprecision is that a limited sample of the population is studied and this source of imprecision is also present in the estimates derived from the standard approach. The second is the added imprecision from the BSM procedure, as the final BSM estimates will depend on the random number sequence used in the simulation-estimation sequence.

The imprecision in parameter estimates from the BSM procedure can be reduced in 2 ways, the parallel and the serial procedure. In the parallel procedure, several BSM estimations are performed and the average of the final estimates from these parallel runs is taken as the final estimate. In the serial procedure, the iteration process of a BSM is extended with *m* additional steps, and the average of the *n* estimates with the lowest value of the convergence criterion is used as the final parameter estimate.

Estimation of contribution of imprecision caused by the BSM and the reduction of imprecision by serial runs were investigated. For this purpose, 20 original data sets were simulated for each of the 3 conditions (1), (2), and (3). Ten of the original data sets were then analyzed 10 times with BSM, resulting in 10 $\Phi_{BSM,single}$ for each data set. Also for the other 10 data sets, 10 BSM estimations were performed, but the average of the 20 estimates with the lowest value of the CC was taken as the final estimate of each run. The estimates from both $\Phi_{BSM,single}$ and $\Phi_{BSM,serial}$, Y_{ij}, were analyzed as independent variables using a linear mixed effects model with 3 parameters:

$$Y_{\text{ij}} = \theta + \theta \cdot \varepsilon_{\text{sample,i}} + \theta \cdot \varepsilon_{\text{BSM,ij}} \quad (9)$$

where the subscript i represents the original data sets, and the subscript j represents each BSM estimate. The θ represents the typical parameter value and $\varepsilon_{sample,i}$ and $\varepsilon_{BSM,ij}$ are zero mean random variables with estimated variances of σ_{sample}^2 and σ_{BSM}^2, respectively. The σ_{sample} is the relative imprecision from sample variability alone or sample imprecision; σ_{BSM} is the imprecision that originates from BSM, either single or serial estimation.

Table 2. Estimated Population Parameters From Standard Approach and Back-Step Method Compared With True Parameter Values*

Condition		$\theta_{bl=1}$	$\theta_{bl=2}$	$\theta_{bl=3}$	θ_{plc}	θ_{dose}	ω^2
(1)	True Value	1.85	-1.85	-1.85	0.483	0.0459	4
	STD	1.84 (1.52, 2.08)	-1.85 (-2.00, -1.69)	-1.84 (-1.95, -1.71)	0.470 (0.228, 0.679)	0.0462 (0.0330, 0.0577)	3.80 (3.13, 4.64)
	BSM	1.84 (1.58, 2.21)	-1.86 (-2.05, -1.67)	-1.86 (-1.98, -1.70)	0.520 (0.238, 0.844)	0.0471 (0.0289, 0.0618)	4.02 (3.40, 4.87)
(2)	True Value	-4.88	-0.548	-1.18	1.55	0.0303	4
	STD	-5.46 (-6.93, -4.53)	-0.571 (-0.698, -0.421)	-1.25 (-1.48,-1.03)	1.56 (0.909, 2.23)	0.0342 (0.00254, 0.0543)	6.05 (3.26, 10.1)
	BSM	-4.84 (-5.61,-4.22)	-0.561 (-0.703, -0.438)	-1.19 (-1.44, -0.905)	1.54 (0.938, 2.34)	0.0326 (0.0155, 0.1512)	4.10 (2.62, 7.58)
(3)	True Value	-11.8	-1.32	-2.96	3.85	0.0717	40
	STD	-14.8 (-16.6, -13.5)	-1.58 (-1.89, -1.19)	-3.59 (-4.71, -2.99)	5.36 (4.26, 6.83)	0.0422 (0.0173, 0.0772)	121 (96.3, 163)
	BSM	-11.6 (-14.4, -9.87)	-1.33 (-1.72, -1.09)	-2.96 (-3.57, -2.39)	3.92 (2.89, 5.26)	0.0725 (0.0212, 0.148)	40.0 (26.5, 69.3)

*STD indicates standard approach; BSM, Back-Step Method; $\theta_{bl=1}$, $\theta_{bl=2}$, and $\theta_{bl=3}$, fixed effect parameters describing the baseline probabilities; θ_{plc} and θ_{dose}, fixed-effect parameters for placebo and drug, respectively; and ω^2, variance of the overall IIV. Results are given for the 3 conditions: (1) nonskewed distribution of response with low interindividual variability (IIV), (2) skewed distribution with low IIV, and (3) skewed distribution with high IIV, as average (range).

Standard Errors for BSM estimates

A nonparametric bootstrap was performed to illustrate how SEs for the parameter estimates from the BSM can be obtained. For this purpose, one original data set was simulated under condition (3) using the nominal parameter values listed in Table 1. Two hundred nonparametric bootstrap data sets were created based on the simulated original data set. Parameters were estimated from the bootstrap data sets, using serial BSM estimations, where the average of the 20 estimates of 350, with the lowest value of the CC, was used as the final estimate. The standard errors of these BSM/bootstrap estimates were then calculated.

In the investigation of imprecision, but not in the investigation of bias, only simulated data with at least one observation in each category were used to assure estimability of the model parameter.

RESULTS

When analyzing the data using the standard approach in NONMEM, the bias in the parameter estimates increased with increasing skewness of the response and with increasing interindividual variability (Table 2 and Figure 2). Simulation of new data based on these estimates showed an increase in overprediction of rare events with increasing bias (Figure 3).

The Back-Step Method performed without bias in all conditions tested (Table 2 and Figure 2). Since Φ_{BSM} were unbiased for all conditions, the predictions of events using BSM estimates were in good agreement with the original data (Figure 3).

The BSM was well suited to estimate the parameters as evidenced by high precision of the estimates from 10 replicate data sets of each condition analyzed using a linear mixed effects model. Even for a single BSM estimation, sample variability dominated over BSM variability (Table 3). However, serial BSM estimates reduced the contribution from BSM even further (Table 3). The σ^2_{BSM} was on average 35% of total variance for a single BSM estimation, whereas for serial BSM estimations, σ^2_{BSM} was decreased to 4% of total variance.

The standard errors, estimated using the BSM in a nonparametric bootstrap, are listed in Table 4. As expected, the relative SE estimates obtained by nonparametric bootstrap were in good agreement with the sample imprecision estimated. None of the 90% confidence intervals, estimated using the 200 bootstrap samples, included zero.

DISCUSSION

In this study, the performance was assessed of a method that reduces the bias in parameter estimates by iteratively finding the estimates that upon simulation result in a data set that closely resembles the observed data. The performance of the Back-Step Method was found to be without appreciable bias under the conditions that were investigated.

The BSM shares many similarities with the posterior predictive check (PPC)[14-16] and with the simulation hypothesis test (SHPT)[17,18] in that it compares some variable value derived from the observed data with a value or distribution of values of the same variable derived from data sets generated from

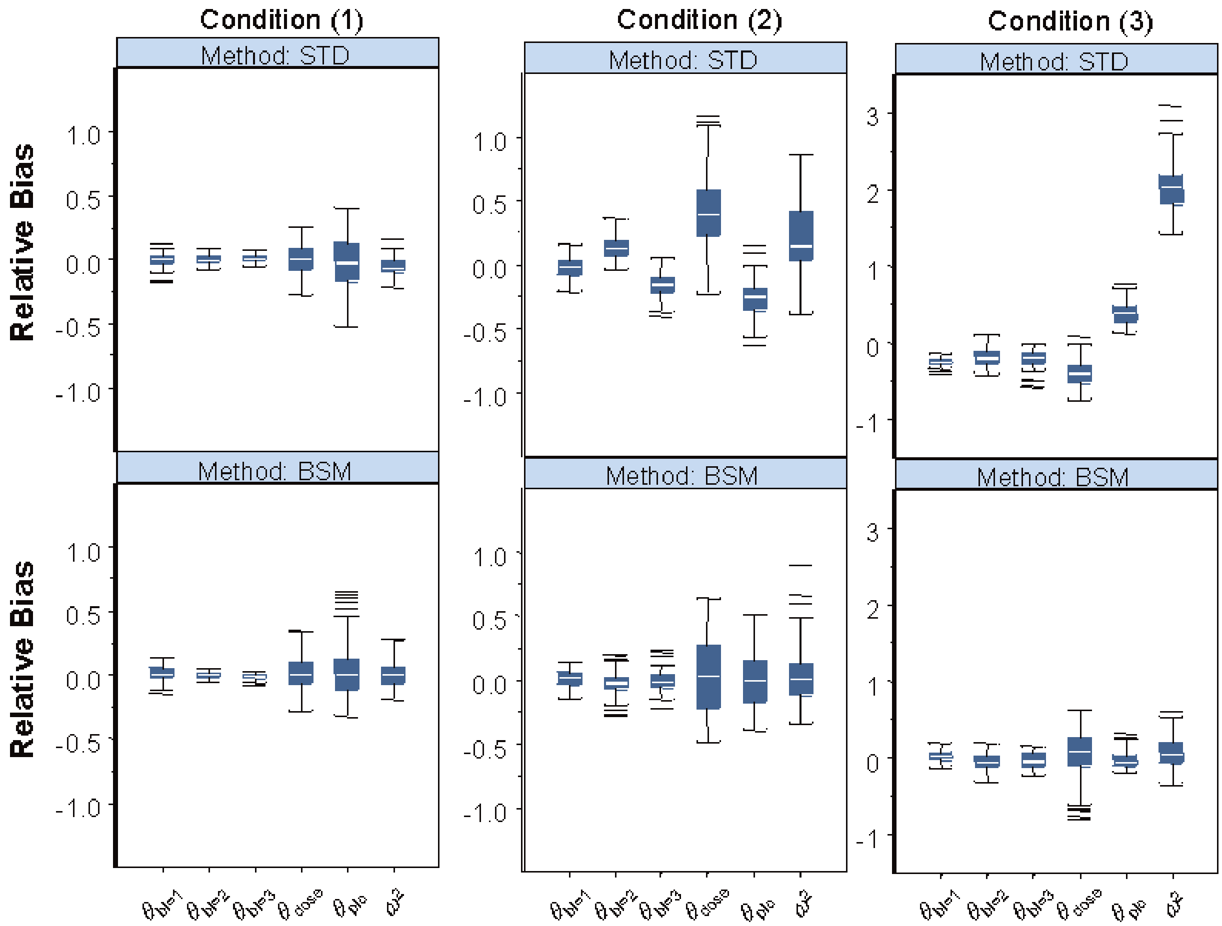

Figure 2. Box and whiskers plots of relative bias based on 100 parameters estimated for the 3 conditions tested: (1) nonskewed distribution of response with low interindividual variability (IIV) (left), (2) skewed distribution with low IIV (middle), and (3) skewed distribution with high IIV (right) using STD in NONMEM (top) and BSM (bottom). $\theta_{bl=1}$, $\theta_{bl=2}$, and $\theta_{bl=3}$ are the fixed-effect parameters describing the baseline probabilities; θ_{plc} and θ_{dose} are the fixed-effect parameters for placebo and drug, respectively; and ω^2 is the variance of the overall IIV. The ends of each box are the 1st and 3rd quartiles, and a line is drawn within the box at the median. Whiskers are drawn from the edges of the box to the most extreme values, provided these values do not extend more than $\pm 3/2\ q$, where q is the interquartile range. Values beyond the whiskers are considered outliers and are indicated by lines.

the model. In both the BSM and the SHPT, a feature of the modeling process is calibrated using model-based simulations. In all 3 procedures, it is believed that comparison between features of data simulated from the model and features of the real data can be used to assess the adequacy of the model. In the PPC, a discrepancy between the model-predicted variable value and the corresponding observed value would result in rejection of the present model and renewed modeling efforts. Similarly, in the BSM, disagreement between the parameter values obtained from the simulated data and the values from the real data will result in a renewed search for parameter estimates. However, whereas in the PPC the modeling-prediction sequence is performed only once or repeated only a few times, it is repeated many times in the BSM. This is one reason that it is advisable to test any model developed by BSM by using a PPC or similar procedure.

The BSM requires a set of initial estimates for the iterative procedure. The results are presented for the case in which these initial estimates are taken to be identical to $\Phi_{est}(1)$. Using the estimates from a naïve pooling of the data and setting the variance parameter to a low value is also an alternative method for choosing initial estimates (Figure 4). For the type of models and data used in this work, the estimates were stable after approximately 20 iterations, no matter what initial estimates were used. This is not necessary true for all types of models and data; as with most iterative procedures, it may be prudent to try different initial estimates to assure that a stable set of final parameter estimates has been found.

To assess the importance of correctly estimating a particular parameter, we have in the CC chosen to divide the absolute difference between $\Phi_{est}(1)$ and $\Phi_{est}(n)$ by the standard error of $\Phi_{est}(1)$. As the skewness of the response distribution

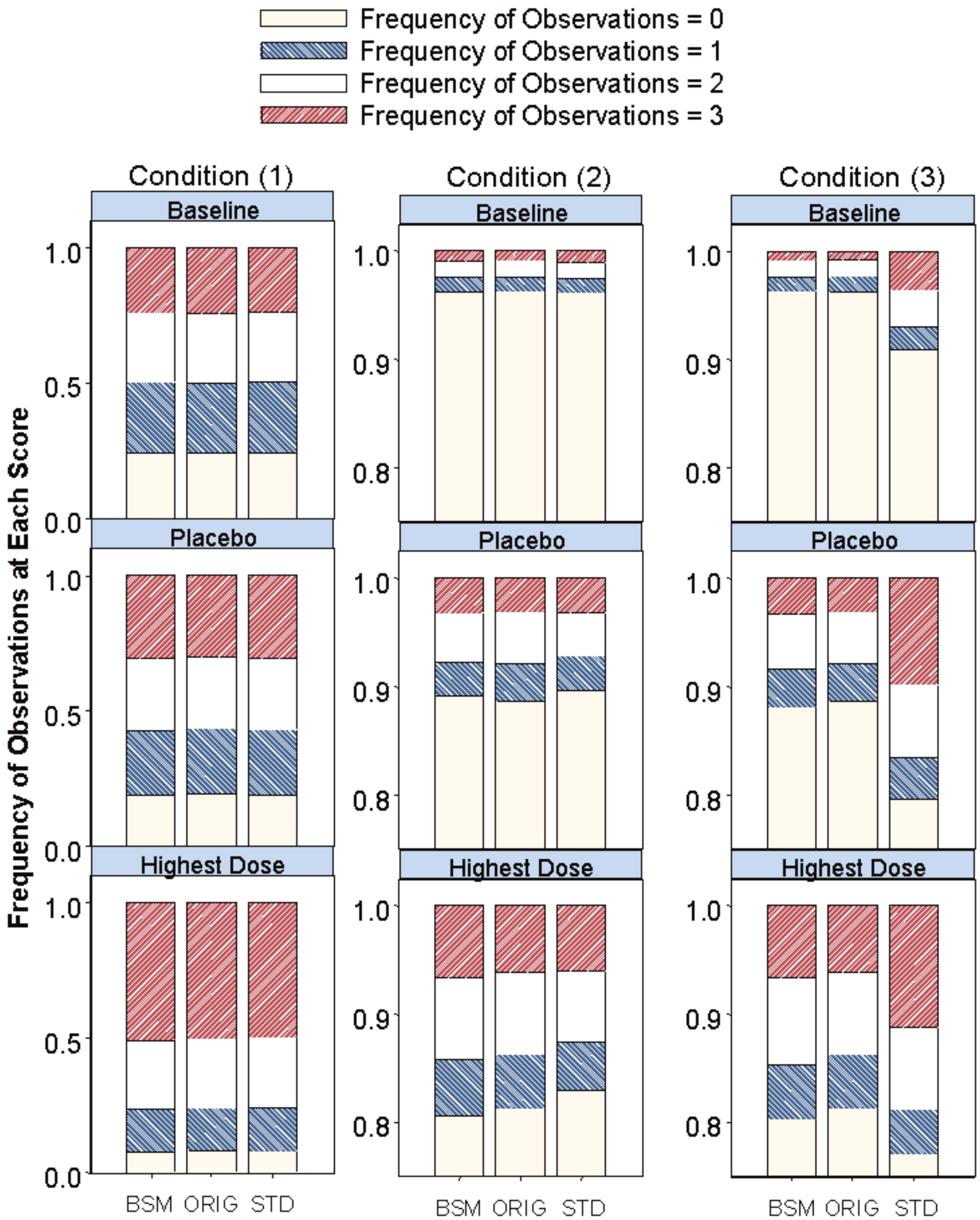

Figure 3. Frequency of events at each score in original data (ORIG) compared with simulated data using estimates from either the STD in NONMEM or BSM for the conditions tested: (1) nonskewed distribution of response with low interindividual variability (IIV), (2) skewed distribution with low IIV, and (3) skewed distribution with high IIV. The simulated frequencies are averages of 100 simulations and are shown at baseline (top), placebo (middle), and the highest dose (bottom). The height of each bar segment in each panel is proportional to the fraction of patients exhibiting 0, 1, 2, and 3 in the response among all patients for the specified condition and for the specified method.

increases, the bias in the parameter estimate increases and the estimate of standard error may not be reliable. However, the SEs are used in the CC as scaling factors, to reassure equal importance of the parameter estimates; only the order of magnitude of the SE estimates is of importance. The model convergence will be affected by bias in the SEs in the sense that the procedure will be prolonged or shortened, resulting in different parameters being chosen as the final than would have been chosen had the SE estimates been unbiased. However, performing several BSM estimations, as we recommend, would reduce the effect on the final estimates and in the end only affect the precision, not the accuracy.

In the present study, the CC was used to shorten the total run time in the cases when the estimated parameters were very close to the original parameters (~10% of the runs). However, an alternative is to continue the simulation-estimation steps for all data sets until the simulation parameter estimates, Φ_{sim}, show a stable pattern and then use the average of the n parameter estimates with the lowest value of the CC as the BSM estimate. An example of such a trace-plot of

Table 3. The sample Imprecision, σ_{sample}, and Imprecision Due to BSM, σ_{BSM}, for Single and Serial Estimations*

	Single											
	$\theta_{bl=1}$		$\theta_{bl=2}$		$\theta_{bl=3}$		θ_{plc}		θ_{dose}		ω^2	
	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}
(1)	0.043	0.041	0.020	0.016	0.020	0.017	0.17	0.14	0.11	0.088	0.11	0.057
(2)	0.049	0.035	0.068	0.058	0.063	0.043	0.12	0.075	0.28	0.47	0.18	0.10
(3)	0.053	0.045	0.11	0.040	0.084	0.041	0.11	0.061	0.26	0.18	0.14	0.10
	Serial											
	$\theta_{bl=1}$		$\theta_{bl=2}$		$\theta_{bl=3}$		θ_{plc}		θ_{dose}		ω^2	
	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}	σ_{sample}	σ_{BSM}
(1)	0.045	0.010	0.020	0.0059	0.019	0.0048	0.19	0.049	0.098	0.023	0.11	0.014
(2)	0.050	0.014	0.072	0.016	0.065	0.014	0.14	0.031	0.44	0.052	0.16	0.045
(3)	0.054	0.0087	0.10	0.016	0.085	0.015	0.11	0.018	0.26	0.042	0.13	0.026

*BSM indicates Back-Step Method; $\theta_{bl=1}$, $\theta_{bl=2}$, and $\theta_{bl=3}$, fixed-effect parameters describing the baseline probabilities; θ_{plc} and θ_{dose}, fixed-effect parameters for placebo and drug, respectively; and ω^2, variance of the overall IIV. Results are given for the 3 conditions: (1) nonskewed distribution of response with low interindividual variability (IIV), (2) skewed distribution with low IIV, and (3) skewed distribution with high IIV.

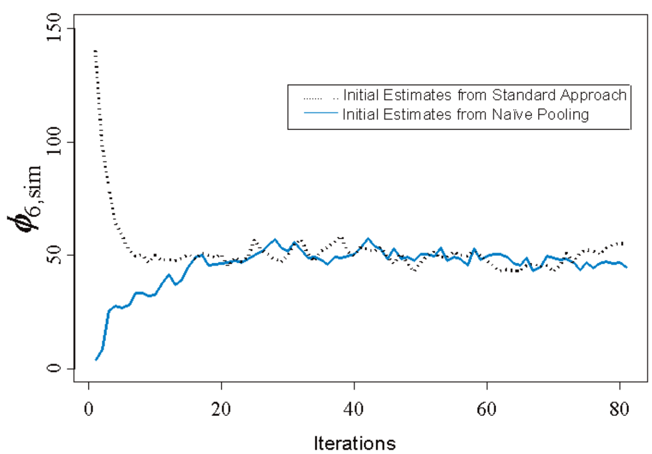

Figure 4. The simulation parameter values for the variance parameter for the overall interindividual variability, ω^2, with 2 different initial estimates. One set of initial parameters was based on the estimates from nonlinear mixed effects modeling and one set was based on the estimates from naïve pooling. In the latter case, the variance parameter was arbitrarily set to be low.

parameter estimates from condition (3) is shown in Figure 5. If this latter alternative is used, it may be advantageous to replace the SE($\Phi_{est}(1)$) in the CC, with the standard deviation from the simulation parameters in the trace-plot to get a more accurate uncertainty estimate. This latter method of assessing the convergence, without using the SE estimate of the first estimation, is also suitable if the covariance step in NONMEM fails.

In addition to the convergence criterion used, the total run time will depend on the run time of a single problem, the difference between the starting and the final estimates and the efficiency of the search algorithm. The run times in the present study were acceptable even though extensive simulations were performed. Nevertheless, run times might become extensive with other types of data or with other models, and search algorithms that are more efficient than the one used in this study might be developed to handle this problem. For estimates that are sensitive to changes in the data structure, it might be more efficient to change one simulation parameter at a time, instead of all at once, as was done in this study. Other search algorithms that might be useful are optimization techniques, such as the steepest descend or the simplex method. However, bias in parameter estimates is in general relatively modest, and the initial parameter estimates should often be not too distant from the final estimates. Indeed, seldom are such severe biases observed as for the skewed ordered categorical data used here.

Bias in parameters has previously been shown in various situations.[2-5] The use of more suitable methods, within or outside of NONMEM, may be precluded because of a lack of knowledge of a more suitable method, lack of access to such a method, prohibitively long run times, or absence of appropriate termination with a more suitable method. Even though we only have investigated the BSM's performance on ordered categorical data, the method may well be applicable to other types of data and other types of models. Since BSM only corrects the parameters affected by the bias, the method can in principle be applied for any model or data. Bias in parameter estimates within NONMEM can sometimes be handled through the centering option.[1] The performance of the centering option in ordered categorical data with a skewed distribution of responses, estimated using the proportional odds model, was investigated by Jönsson et al.[10] The centering option decreased the bias considerably but did not solve the problem. Thus, the BSM may be an attractive alternative in these situations.

Using BSM when the bias arises from model misspecification and not from estimation method approximations will not

Table 4. Relative Standard Error of a Nonparametric Bootstrap, Estimated Using Back-Step Method*

$\theta_{bl=1}$	$\theta_{bl=2}$	$\theta_{bl=3}$	θ_{dose}	θ_{plc}	ω^2
0.070	0.080	0.070	0.11	0.31	0.14

*The original data were simulated under condition (3), data with a skewed distribution and high IIV. $\theta_{bl=1}$, $\theta_{bl=2}$ and $\theta_{bl=3}$ are the fixed-effect parameters describing the baseline probabilities; θ_{plc} and θ_{dose} are the fixed effect parameters for placebo and drug, respectively; and ω^2 is the variance of the overall IIV.

improve the predictions. To assess whether the bias can be corrected using BSM, data can be repeatedly simulated using the model and then analyzed using the same model to obtain a set of estimated parameters for each simulated data set.[19] If the estimated parameters are not biased compared with the ones used for simulation, using BSM will not improve the fit.

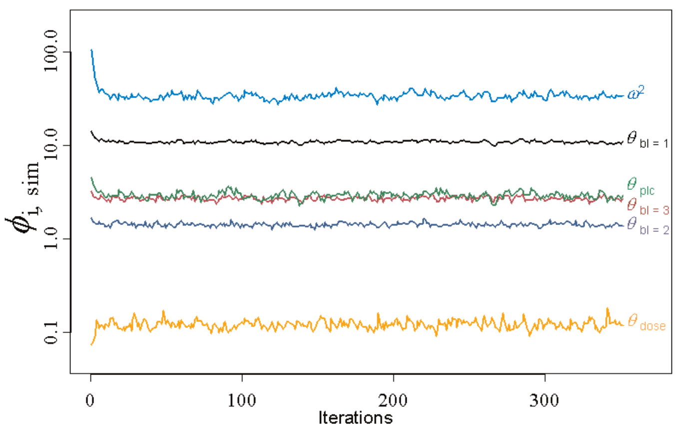

Figure 5. Simulation parameter values during a BSM run. $\theta_{bl=1}$, $\theta_{bl=2}$, and $\theta_{bl=3}$ are the fixed-effect parameters describing the baseline probabilities; θ_{plc} and θ_{dose} are the fixed-effect parameters for placebo and drug, respectively; and ω^2 is the variance of the overall interindividual variability (IIV).

With BSM, the objective function value relates to a simulated data set and can therefore not be used for model discrimination. As an alternative, the confidence intervals of the parameters could be used for decisions of whether to include or not include a particular parameter-covariate relation in a model. However, the standard errors provided by NONMEM for the original data set should be treated with considerable caution as they represent imprecision estimates from a flawed fit. Standard errors could preferably be obtained by nonparametric bootstrap of the original data set and estimate parameters for each bootstrap sample using BSM, as was done for one example data set in this work. In that example, none of the ranges of the estimated parameters included zero, thus including dose as a significant parameter was evident. With 200 bootstrap samples, the true SEs of the parameters are estimated, but to assess the true 95% confidence intervals, it is recommended that at least 2000 nonparametric bootstrap samples are performed.[20] Inclusion of covariates, based on the SE estimates, assumes that the distribution of the parameter estimates is symmetrical, and a visual check, to test if this assumption is valid, should be performed.

As bootstraps of the BSM may be run-time intensive, a relatively condensed model building would be advantageous. At least for ordered categorical data, such condensed model building may fortunately be feasible, as the structural models generally are simple, usually having only a single random effect, and candidate covariate models may be identified through regression against empirical Bayes etas or from more exhaustive methods using naïve pooling. Yano et al[21] have shown that for dichotomous data there is no advantage in using nonlinear mixed effects modeling over naïve pooling, when estimating fixed-effect parameters. In the case of skewed ordered categorical data estimated in NONMEM, naïve pooling could probably be used for forward covariate model building, and BSM could be used for determining the error magnitude and structure with nonlinear mixed effects modeling. Naïve pooling suffers from the erroneous assumption that all observations are independent and therefore the standard likelihood ratio test cannot be used, but this problem can to a large extent be alleviated by a randomization test.[22]

In conclusion, the BSM is an alternative method for estimation and simulation when the standard approach in NONMEM produces biased estimates. For skewed ordered categorical data, BSM demonstrated good accuracy even when the standard method yielded strongly biased parameter estimates.

ACKNOWLEDGEMENTS

This study was supported in part by Pfizer Central Research, Sandwich, UK (Maria C. Kjellsson) and by AstraZeneca Research and Development, Södertälje, Sweden (Siv Jönsson).

REFERENCES

1. SL Beal, LB Sheiner. NONMEM users guides (I-VIII). Hanover, Maryland: GloboMax: 1989-1998.

2. White DB, Walawander CA, Tung Y, et al. An evaluation of point and interval estimates in population pharmacokinetics using NONMEM analysis. *J Pharmacokinet Biopharm*. 1991;19:87-112.

3. Bennett JE, Wakefield JC. A comparison of a Bayesian population method with two methods as implemented in commercially available software. *J Pharmacokinet Biopharm*. 1996;24:403-432.

4. Wakefield J, Walker S. A population approach to initial dose selection. *Stat Med*. 1997;16:1135-1149.

5. EN Jonsson. JR Wade, MO Karlsson. Nonlinearity detection: advantages of nonlinear mixed-effects modeling. *AAPS PharmSci*. 2000;2(3):E32. Available from: http://www.aapspharmsci.org/.

6. Gupta SK, Sathyan G, Lindemulder EA, et al. Quantitative characterization of therapeutic index: Application of mixed-effects modeling to evaluate oxybutynin dose-efficacy and dose-side effect relationships. *Clin Pharmacol Ther*. 1999;65:672-684.

7. Mandema JW, Stanski DR. Population pharmacodynamic model for ketorolac analgesia. *Clin Pharmacol Ther*. 1996;60:619-635.

8. Mould DR, Holford NH, Schellens JH, et al. Population pharmacokinetic and adverse event analysis of topotecan in patients with solid tumors. *Clin Pharmacol Ther*. 2002;71:334-348.

9. Sheiner LB. A new approach to the analysis of analgesic drug trials, illustrated with bromfenac data. *Clin Pharmacol Ther*. 1994;56:309-322.

10. Jönsson S. Estimation of dosing strategies for individualisation [dissertation]. Uppsala, Sweden: Acta Universitatis Upsaliensis; 2004. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-4255.

11. Kowalski KG, McFadyen L, Hutmatcher MM, et al. A Two-Part Mixture Model for Longitudinal Adverse Event Severity Data. *J Pharmacokinet Pharmacodyn*. 2003;30:315-336.

12. Olsen MK, Schafer JL. A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data. *J Am Stat Assoc*. 2001;96:730-745.

13. Hietaniemi J. Comprehensive Perl Archive Network. http://www.cpan.org (October 2003).

14. Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. London, UK: Chapman & Hall; 1995.

15. Girard P, Blaschke TF, Kastrissios H, et al. A Markov mixed effect regression model for drug compliance. *Stat Med*. 1998;17:2313-2333.

16. Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J Pharmacokinet Pharmacodyn*. 2001;28:171-192.

17. Beal SL. Commentary on significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn*. 2002;29:403-410;discussion 411-412.

18. Gisleskog PO, Karlsson MO, Beal SL. Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn*. 2002;29:473-505.

19. Karlsson MO, Jonsson EN, Wiltse CG, et al. Assumption testing in population pharmacokinetic models: illustrated with an analysis of moxonidine data from congestive heart failure patients. *J Pharmacokinet Biopharm*. 1998;26:207-246.

20. Efron B. Better Bootstrap Confidence Intervals. *J Am Stat Assoc*. 1987;82:171-185.

21. Yano I, Beal SL, Sheiner LB. The Need For Mixed-Effects Modeling with Population Dichotomous Data. *J Pharmacokinet Pharmacodyn*. 2001;28:389-412.

22. Wählby U, Jonsson EN, Karlsson MO. Assessment of Actual Significance Levels for Covariate Effects in NONMEM. *J Pharmacokinet Pharmacodyn*. 2001;28:231-252.